

# Approximation scheme based on effective interactions for stochastic gene regulation

Jun Ohkubo\*

*Graduate School of Informatics, Kyoto University,  
36-1, Yoshida Hon-machi, Sakyo-ku, Kyoto-shi, Kyoto 606-8501, Japan*

(Dated: September 17, 2010)

Since gene regulatory systems contain sometimes only a small number of molecules, these systems are not described well by macroscopic rate equations; a master equation approach is needed for such cases. We develop an approximation scheme for dealing with the stochasticity of the gene regulatory systems. Using an effective interaction concept, original master equations can be reduced to simpler master equations, which can be solved analytically. We apply the approximation scheme to self-regulating systems with monomer or dimer interactions, and a two-gene system with an exclusive switch. The approximation scheme can recover bistability of the exclusive switch adequately.

## I. INTRODUCTION

Recently, stochastic nature in small systems has attracted many attentions [1–3]. One of the interesting examples of the stochasticity is a gene regulatory system; it has been known experimentally that the gene regulatory systems show various phenomena caused by intrinsic noise [4, 5]. The gene regulatory systems basically consists of genes, RNAs, and proteins. The genes could sometimes be activated or repressed by regulatory proteins known as transcription factors. The number of regulatory proteins is sometimes very small, and there are large fluctuations. From a theoretical point of view, the gene regulatory systems have been studied a lot using Monte Carlo simulations (e.g., [6, 7]). In addition, in order to gain insights into mechanisms or functions of the gene regulatory systems, many analytical studies have been done [8–18]. For example, if we consider a self-regulating gene with monomer binding interactions, an exact solution has been already known [12]. When one considers more complicated systems, some approximations are needed. Such approximations have also been developed; Fokker-Planck or Langevin equation approach [8–10], a variational approach [11, 14, 15], and self-consistent proteomic field approximation [13].

A gene regulatory system with only two genes and feedback mechanisms has been studied a lot because it plays an important role as a genetic switch; two distinct stable states emerges, and they could be switched either spontaneously or by external signals. In mathematical description for the gene regulatory systems, the RNAs are sometimes neglected for simplicity, and only genes and regulatory proteins are considered. When we construct a macroscopic rate equation, in which fluctuations in protein copy numbers or gene expression states are neglected, the analysis for the rate equation tells us the following facts: A system with two mutually repressing genes shows a bistability, and cooperative binding of regulatory proteins is important for making the bistability [19, 20]. Here, the cooperative binding means that com-

binations of two or more proteins need to activate or repress genes. The macroscopic rate equation gives multiple stable solutions, and each solution corresponds to a stable state of the gene regulatory systems, which causes the bistability. Hence, for the cooperative binding cases, it may be enough to use the macroscopic treatments in order to investigate qualitative behavior of the bistability. However, other studies have shown that a so-called exclusive switch shows a bistability even when the macroscopic rate equations have only one solution [21–23]. Although the bistability has been confirmed numerically using Monte Carlo simulations, no exact or approximated analytical treatment has been proposed yet, to the best of our knowledge.

In the present paper, we develop a new approximation scheme for gene regulatory systems. In the approximation scheme, there is no need to use continuous description such as Fokker-Planck or Langevin equations, and hence the smallness or discrete properties of the system are not neglected. The basic idea of the approximation is similar to the “self-consistent proteomic field approximation” developed by Walczak *et al.* [13]. In the self-consistent proteomic field approximation, a joint probability for all genes is approximated as a product of probability distributions for each gene, and then the interactions between genes and regulatory proteins can be evaluated ‘exactly’ in this approximation. In [13], only toggle switches, which consist of two genes, have been studied; as denoted in the discussions in [13], further approximation would be needed for self-regulating systems. We here extend the concept of [13], and develop a more applicable approximation scheme; the interactions between genes and regulatory proteins are approximated firstly, and an effective interaction is introduced. The new approximation scheme would be useful to treat more complicated cases, such as the exclusive switch. The new approximation scheme enables us to give analytical expressions for probability distributions of the numbers of proteins, without loss of the discreteness property of the system. The effective interactions are estimated self-consistently. We will demonstrate the usefulness of the approximation scheme by using self-regulating systems and the exclusive switch without cooperative interactions.

The present paper is constructed as follows. In Sec. II,

---

\*Email address: ohkubo@i.kyoto-u.ac.jp

we give a brief review of a stochastic model for gene regulation. In Sec. III, self-regulating systems are studied. Section III C gives one of the important results in the present paper, in which our approximation scheme is proposed. The proposed approximation scheme is applied to the exclusive switch in Sec. VI. Section V is concluding remarks.

## II. STOCHASTIC MODEL FOR GENE REGULATION

We here briefly review the basic biology of genetic regulatory system and a simplified stochastic model, for readers' convenience.

A gene regulatory system consists of many components, such as genes, RNAs, and proteins. The transcription of a gene is initiated by a binding of RNA polymerase to a promoter site of the gene in the DNA. The binding of regulatory proteins (or molecules), so-called transcription factors, can sometimes regulate the transcription initiation. These regulatory proteins bind to own target operator sites, and they sometimes act as repressors (which repress the transcription) or activators (which enhance the transcription) of the transcription. When the RNA polymerase binds to a gene, the gene sequence is copied into a messenger RNA (mRNA), and the mRNA is translated into a protein molecules by a ribosome enzyme complex. The produced proteins are important to determine the phenotypic behavior of the cell. In addition, regulation of transcription is one important way of controlling the phenotypic behavior, and sometimes the produced proteins can become regulatory signals for genes.

Although all of the above reactions would be important for the gene regulatory systems, the mRNA is sometimes neglected in stochastic modelling for simplicity. That is, the translation from mRNAs to proteins are straightforward, and then we assume that an activated gene directly increases the number of proteins. In addition, we consider that a repressed gene cannot produce any proteins, which makes analytical treatments much simpler [12].

In the present paper, all regulatory proteins act as repressors. If regulatory proteins are not binding to a gene, then we call a state of the gene as 'ON' state; if not, the gene is in 'OFF' state. A gene in OFF state cannot produce any proteins, as we assumed above.

## III. SELF-REGULATING SYSTEM

In this section, we will explain a new approximation scheme using a simplest model, i.e., a self-regulating system. Exact solutions for the self-regulating system with monomer interactions have already been known. After reviewing the exact solutions, we will propose a new approximation scheme. The new approximation scheme will

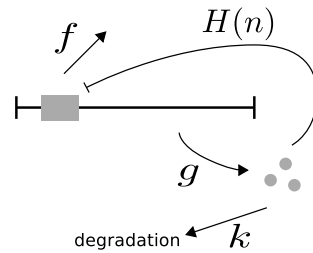


FIG. 1: A schematic illustration of the self-regulating gene.

be applied to the self-regulating systems with monomer and dimer interactions, respectively.

### A. Model

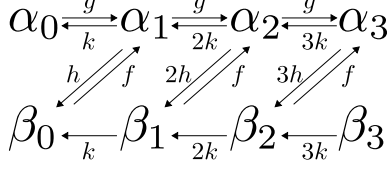
At first, we give a brief explanation for a self-regulating system. In the self-regulating system, there is only one gene, and it produces proteins. The produced proteins are considered as regulatory proteins for the gene, and the regulation is a repressed one. In this sense, there is a self-regulation mechanism. Figure 1 shows the self-regulating system. When the gene is in ON state, it produces proteins with rate  $g$ . The degradation rate of the regulatory proteins is  $k$ . The regulatory proteins can bind the gene with rate function  $H(n)$ , where  $n$  is the number of 'free' regulatory proteins. The function  $H(n)$  can be a complicated function of the regulatory proteins; e.g.,  $H(n) = hn$  for monomer interactions, and  $H(n) = hn(n-1)/2$  for dimer interactions, where  $h$  is a rate for the binding.  $f$  is the rate with which the regulatory protein is released from the repressor site of the gene.

### B. Exact solution for monomer interactions

We here consider a simplest interaction case, i.e., a monomer interaction case. Hence,  $H(n)$  in Fig. 1 is written as  $hn$ , as discussed in Sec. III A. For the monomer interaction cases, exact solutions have already been known [6, 12]. In order to compare our approximation scheme, which will be proposed in Sec. III C, we here briefly review the exact solutions.

In order to make analytical treatments simpler, one assumption should be included [6]; i.e., one of the proteins in ON state is inert, and then the protein cannot be degraded or repress to the gene. Hence, there are a little difference between usual stochastic simulations and this analytical treatment. However, it has already been discussed that this assumption alter only for lower numbers of proteins, and actually it gives quantitatively good results [6]. Hence, we here employ this assumption. Figure 2 shows the transition scheme for the usual stochastic simulations and the analytical treatment.  $\alpha_n$  and  $\beta_n$  correspond to probabilities with which there are  $n$

### Stochastic simulation



### Analytic solution

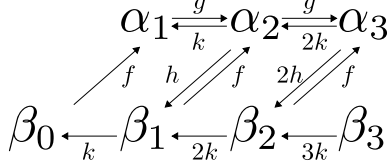


FIG. 2: Transition scheme for simulations and analytical calculations for monomer interaction cases.  $\alpha_n$  and  $\beta_n$  correspond to probabilities with which there are  $n$  regulatory proteins for ON and OFF states, respectively. In the analytical calculations, one of the proteins is considered as an inert one when the gene is in ON state, and the inert protein is also included in  $\alpha_n$ ; the number of ‘free’ regulatory proteins in ON state is  $n - 1$ .

regulatory proteins for ON and OFF states, respectively. In the usual stochastic simulations, the degradation rate of the proteins, i.e., the change from  $\alpha_n$  to  $\alpha_{n-1}$ , is proportional to the number of proteins,  $n$ . In contrast, the above assumption means that the rate from  $\alpha_n$  to  $\alpha_{n-1}$  in the analytical treatment is proportional to  $n - 1$ , not to  $n$ .

In this assumption, the master equations are given as

$$\frac{d\alpha'_n}{dt} = g[\alpha'_{n-1} - \alpha'_n] + k[(n+1)\alpha'_{n+1} - n\alpha'_n] - hn\alpha'_n + f\beta_n, \quad (1)$$

$$\frac{d\beta_n}{dt} = +k[(n+1)\beta_{n+1} - n\beta_n] + hn\alpha'_n - f\beta_n, \quad (2)$$

where  $\alpha'_n$  is a probability with which there is  $n$  ‘free’ regulatory proteins for ON state;  $\alpha_{n+1} \equiv \alpha'_n$ . Note that the inert protein is not a ‘free’ regulatory protein, and  $\alpha'_n$  does not include the inert protein.

In order to solve eqs. (1) and (2), it is useful to define the following generating functions;

$$\alpha'(z) = \sum_{n=0}^{\infty} \alpha'_n z^n, \quad (3)$$

$$\beta(z) = \sum_{n=0}^{\infty} \beta_n z^n. \quad (4)$$

Using the generating functions, various information about the self-regulating system can be obtained. For

example, the number of ‘free’ regulatory proteins is

$$\alpha'_n = \frac{\partial^n}{\partial z^n} \alpha'(z) \Big|_{z=0}. \quad (5)$$

The probability with which the gene is in ON state is given as  $\alpha'(1)$ ; the number of total proteins in the system is given as

$$\langle n \rangle = \frac{\partial \alpha'(z)}{\partial z} \Big|_{z=1} + 1 \times \alpha'(1) + \frac{\partial \beta(z)}{\partial z} \Big|_{z=1}, \quad (6)$$

where the second term in r.h.s means a contribution from the inert protein in ON state.

Eqs. (1) and (2) can be rewritten as two differential equations in terms of the generating functions;

$$\frac{\partial \alpha(z)}{\partial t} = (z-1) \left[ g\alpha'(z) - k \frac{\partial \alpha'}{\partial z} \right] - hz \frac{\partial \alpha'}{\partial z} + f\beta(z), \quad (7)$$

$$\frac{\partial \beta(z)}{\partial t} = -(z-1)k \frac{\partial \beta}{\partial z} + hz \frac{\partial \alpha'}{\partial z} - f\beta(z). \quad (8)$$

After some calculations, ‘stationary’ solutions for the generating functions are obtained as follows [6, 12]:

$$\alpha'^{\text{ex}}(z) = A^{\text{ex}} F(1 + a^{\text{ex}}, 1 + b^{\text{ex}}, N^{\text{ex}}(z - z_0^{\text{ex}})), \quad (9)$$

$$\beta^{\text{ex}}(z) = -\frac{1}{f}(z-1) \left[ g\alpha'^{\text{ex}}(z) - k \frac{\partial \alpha'^{\text{ex}}(z)}{\partial z} \right] + \frac{h}{f} z \frac{\partial \alpha'^{\text{ex}}(z)}{\partial z}, \quad (10)$$

where the super-script ‘ex’ means ‘exact solutions’, and

$$z_0^{\text{ex}} = \frac{k}{k+h}, \quad N^{\text{ex}} = \frac{g}{k+h}, \\ a^{\text{ex}} = \frac{f}{k}, \quad b^{\text{ex}} = \frac{f}{k+h} + (1 - z_0^{\text{ex}})N^{\text{ex}}. \quad (11)$$

$A^{\text{ex}}$  is the normalization constant, which is determined as  $\alpha'^{\text{ex}}(1) + \beta^{\text{ex}}(1) = 1$ .  $F(p, q, r)$  is the Kummer confluent hypergeometric function,

$$F(p, q, r) \equiv \sum_{n=0}^{\infty} \frac{(p)_n}{(q)_n} \frac{r^n}{n!}, \quad (12)$$

where  $(p)_n = p(p+1)(p+2)\cdots(p+n-1)$ .

Details of the characteristics of the exact solution are written in [6]. For example, the probability distributions for the numbers of ‘free’ regulatory proteins are

$$\alpha'_n{}^{\text{ex}} = \frac{A^{\text{ex}}}{n!} (N^{\text{ex}})^n \frac{(1 + a^{\text{ex}})_n}{(1 + b^{\text{ex}})_n} \\ \times F(1 + a^{\text{ex}} + n, 1 + b^{\text{ex}} + n, -N^{\text{ex}}z_0^{\text{ex}}), \quad (13) \\ \beta_n{}^{\text{ex}} = \frac{A^{\text{ex}}}{f} \left[ ((1 - z_0^{\text{ex}})N^{\text{ex}} - b^{\text{ex}}) \frac{(N^{\text{ex}})^n}{n!} \right. \\ \times \frac{(1 + a^{\text{ex}})_n}{(1 + b^{\text{ex}})_n} F(1 + a^{\text{ex}} + n, 1 + b^{\text{ex}} + n, -N^{\text{ex}}z_0^{\text{ex}}) \\ \left. + b^{\text{ex}} \frac{(N^{\text{ex}})^n}{n!} \frac{(a^{\text{ex}})_n}{(b^{\text{ex}})_n} F(a^{\text{ex}} + n, b^{\text{ex}} + n, -N^{\text{ex}}z_0^{\text{ex}}) \right], \quad (14)$$

where  $\tilde{f} = f/(k + h)$ .

### C. Approximation scheme

In Sec. IIIB, we analyzed the self-regulating system with monomer interactions. In the system, the interaction factor,  $H(n)$ , is given as  $hn$ , and actually this simple form of the interaction enables us to obtain the exact solutions. If we consider different types of interactions, such as dimer ones, exact solutions have not been known yet.

Here, we propose a new approximation scheme. The key of the approximation scheme is to use “an effective interaction”. Although the new approximation scheme is similar to the self-consistent proteomic field approximation in [13], the new one is more applicable; it is applicable even to the self-regulating systems or exclusive switch cases, as shown later. The effective interaction means that the interaction factor in Fig. 1,  $H(n)$ , is replaced as a scalar value; i.e.,  $\mathcal{H}(n) = \tilde{h}$ . Here, the effective interaction  $\tilde{h}$  is not a function of the regulatory proteins. Hence, the master equation for this approximated system is written as follows:

$$\frac{d\alpha'_n}{dt} = g[\alpha'_{n-1} - \alpha'_n] + k[(n+1)\alpha'_{n+1} - n\alpha'_n] - \tilde{h}\alpha'_n + f\beta_n, \quad (15)$$

$$\frac{d\beta_n}{dt} = +k[(n+1)\beta_{n+1} - n\beta_n] + \tilde{h}\alpha'_n - f\beta_n. \quad (16)$$

Because the interaction factor  $H(n) = \tilde{h}$  has a simple form, the analytic solution can be easily calculated by using the generating function approach. Putting the left-hand sides of Eqs. (15) and (16) as zero and rewriting Eqs. (15) and (16) in terms of the generating functions  $\alpha'(z)$  and  $\beta(z)$ , stationary solutions for the generating functions are obtained as follows:

$$\alpha'(z) = A F(a, b, N(z-1)), \quad (17)$$

$$\beta(z) = \left(1 + \frac{h}{f}\right) A F(a-1, b-1, N(z-1)) - \alpha(z), \quad (18)$$

where  $A = f/(f + \tilde{h})$  and

$$N = \frac{g}{k}, \quad a = 1 + \frac{f}{k}, \quad b = 1 + \frac{f + \tilde{h}}{k}. \quad (19)$$

A remaining task is to determine the effective interaction  $\tilde{h}$ . For the self-regulating system with monomer interactions, the binding of the regulatory proteins occurs only when the system is in ON state. Hence, the number of proteins, which can be attached to the binding site, should be equal to the number of free proteins for ON state.

According to the following discussions, we here set the effective interaction  $\tilde{h}$  as

$$\tilde{h} = h\langle n \rangle_{\alpha'}, \quad (20)$$

where  $\langle n \rangle_{\alpha'}$  is the expectation of the number of free regulatory proteins under a condition that the gene is in ON state (conditional expectation). Because it is possible to evaluate the conditional expectation using the generating function (Eq. (17)) as

$$\langle n \rangle_{\alpha'} \equiv \frac{1}{\alpha'(1)} \left. \frac{\partial}{\partial z} \alpha'(z) \right|_{z=1} = \frac{g(k+f)}{k(f+f+\tilde{h})}, \quad (21)$$

we obtain the following self-consistent equation by inserting Eq. (20);

$$h\langle n \rangle_{\alpha'} = h \frac{g(k+f)}{k(f+f+h\langle n \rangle_{\alpha'})}. \quad (22)$$

Solving Eq. (22), we finally obtain

$$h\langle n \rangle_{\alpha'} = \frac{-(k^2 + kf) + \sqrt{(k^2 + kf)^2 + 3khg(k+f)}}{2k}. \quad (23)$$

Once the effective interaction  $\tilde{h}$  is determined, all statistical properties related to the number of regulatory proteins are evaluated from the generating functions (Eqs. (17) and (18)). For example, the probability distributions for the numbers of free proteins are

$$\alpha'_n = AN^n \frac{(a)_n}{(b)_n} F(a+n, b+n, -N), \quad (24)$$

$$\beta_n = \left(1 + \frac{h}{f}\right) AN^n \frac{(a-1)_n}{(b-1)_n} \times F(a-1+n, b-1+n, -N) - \alpha'_n. \quad (25)$$

### D. Results for monomer interactions

For monomer interaction cases, the exact solutions are obtained. Hence, we here compare the exact results and approximate results obtained by the approximation scheme.

For the comparison, we here introduce rescaled parameters as follows [6]:

$$\omega = \frac{f}{k}, \quad X_{\text{eq}} = \frac{f}{h}, \quad X_{\text{ad}} = \frac{g}{2k}, \quad (26)$$

and for simplicity, we set  $k = 1$  in all numerical evaluations. These rescaled parameters are helpful to understand properties of the genetic switch. The parameter  $X_{\text{ad}}$  characterizes the synthesis/degradation processes, and large  $X_{\text{ad}}$  would give a large average number of proteins.  $X_{\text{eq}}$  is related to the equilibrium constant of the binding/unbinding process. Finally,  $\omega$  is a parameter called “adiabaticity parameter”.  $\omega$  measures how rapidly

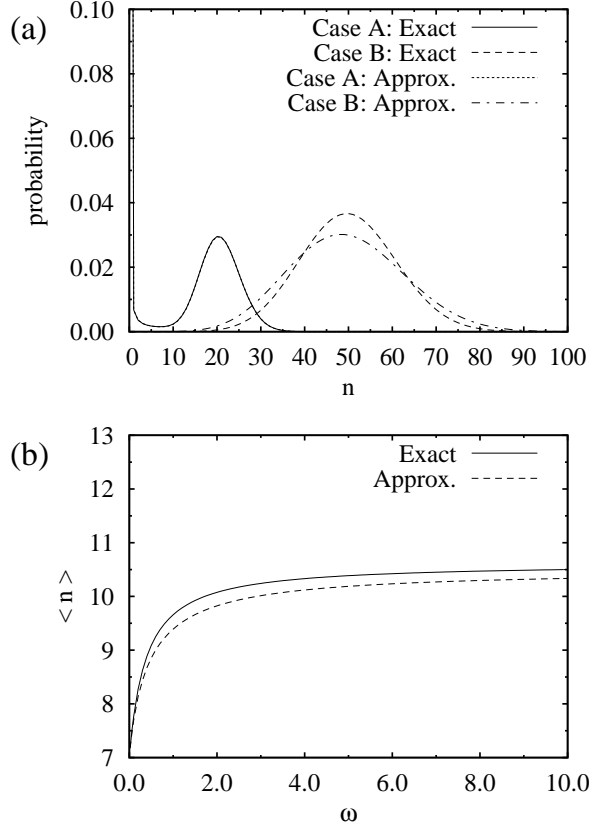


FIG. 3: Comparison between results from the exact solutions and the approximation scheme. (a) The probability distributions of the number of proteins. Case A:  $X_{\text{eq}} = 10.0$ ,  $X_{\text{ad}} = 10.0$ ,  $\omega = 0.01$ . Case B:  $X_{\text{eq}} = 50.0$ ,  $X_{\text{ad}} = 50.0$ ,  $\omega = 10.0$ . For Case A, it is difficult to see the difference between the exact solution and approximate solution. (b) The average number of proteins.  $X_{\text{eq}} = 10.0$  and  $X_{\text{ad}} = 10.0$ , and only the value of  $\omega$  was changed.

the gene can equilibrate in a gene state. If  $\omega$  is small, the synthesis/degradation behaves almost like an independent birth and death process, and there would be two peaks corresponding to the binding/unbinding states, respectively. For details of these parameters, e.g., see [6].

Firstly, the probability distributions of the number of protein were compared. Figure 3(a) shows the results. Here, we performed two cases: In Case A, we set  $X_{\text{eq}} = 10.0$ ,  $X_{\text{ad}} = 10.0$ ,  $\omega = 0.01$ ; in Case B,  $X_{\text{eq}} = 50.0$ ,  $X_{\text{ad}} = 50.0$ ,  $\omega = 10.0$ . For Case A, the exact solution and the approximate one give a good agreement, and it is difficult to see the difference. Although there are quantitative differences between the exact and approximate solutions for Case B, the approximation scheme gives qualitatively good result despite of the rough approximation. Figure 3(b) shows the average number of proteins (Eq. (6)) for various values of  $\omega$  when  $X_{\text{eq}} = 10.0$  and  $X_{\text{ad}} = 10.0$ . If  $\omega$  is small, the approximation scheme gives quantitatively good results. Even in the large  $\omega$

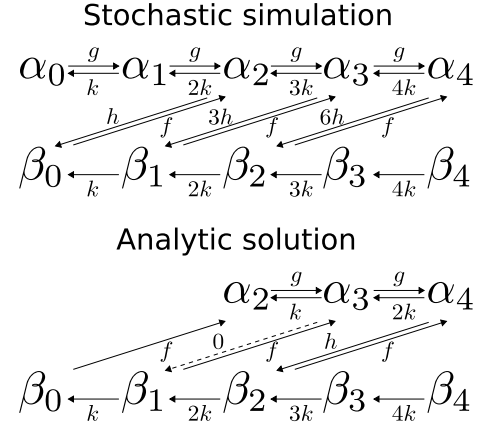


FIG. 4: Transition scheme for simulations and analytical calculations for dimer interaction cases.

case, the difference between the exact and approximate results are less than 1.

### E. Results for dimer interactions

As a second example, a self-regulating system with dimer interactions are considered. In this case, the transition scheme for analytical calculations are different from the monomer interaction cases; see Fig. 4. In this case, the master equations are the same as Eqs. (15) and (16), but the effective interaction should be set as

$$\tilde{h} = h \frac{\langle n(n-1) \rangle_{\alpha'}}{2}, \quad (27)$$

and we should interpret  $\alpha'_n$  as  $\alpha_{n+2} = \alpha'_n$ .

Using the similar procedure written in Sec. III C, the effective interaction  $\tilde{h}$  is obtained by solving the following self-consistent equation:

$$h \frac{\langle n(n-1) \rangle_{\alpha'}}{2} = \frac{h}{2} \frac{1}{\alpha'(1)} \left. \frac{\partial^2}{\partial z^2} \alpha'(z) \right|_{z=0}. \quad (28)$$

Since it is a little complicated task to obtain the analytical expression for the effective interaction  $\tilde{h}$ , we numerically solved the self-consistent equation (Eq. (28)).

For the dimer interaction cases, we have not obtained any exact solution. Hence, results of the approximation scheme were compared with those of the Monte Carlo simulations.

Figure 5 shows the probability distributions. In Case A, we used the following rescaled parameters:  $X_{\text{eq}} = 10.0$ ,  $X_{\text{ad}} = 10.0$ ,  $\omega = 0.01$ ; in Case B,  $X_{\text{eq}} = 50.0$ ,  $X_{\text{ad}} = 50.0$ ,  $\omega = 10.0$ . The numbers of the Monte Carlo steps are over  $10^7$  for Case A, and  $10^8$  for Case B. Obviously, the approximation scheme gives qualitatively good results; although the shapes of the distributions and the positions of peaks are slightly different, the number of peaks are the same as the Monte Carlo results. In addition, the average number of proteins are almost the same

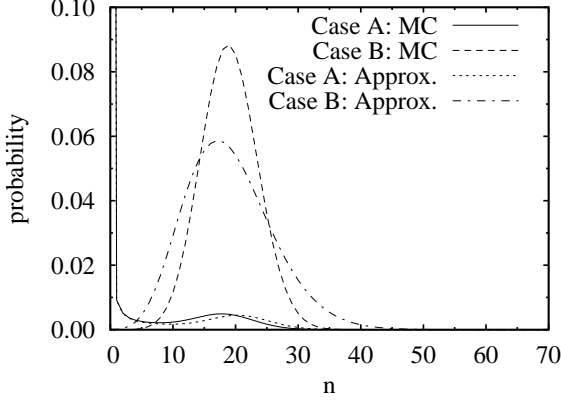


FIG. 5: The probability distributions of the number of proteins. Case A:  $X_{eq} = 10.0$ ,  $X_{ad} = 10.0$ ,  $\omega = 0.01$ . Case B:  $X_{eq} = 50.0$ ,  $X_{ad} = 50.0$ ,  $\omega = 10.0$ .

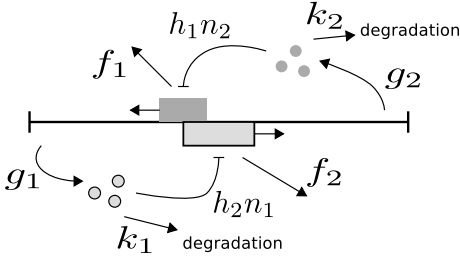


FIG. 6: A schematic illustration of the exclusive switch.

as the Monte Carlo results: for Case A,  $\langle n \rangle = 1.2$  in the Monte Carlo simulation and  $\langle n \rangle = 1.3$  in the approximation scheme; for Case B,  $\langle n \rangle = 19.0$  in the Monte Carlo simulation and  $\langle n \rangle = 19.0$  in the approximation scheme.

#### IV. EXCLUSIVE SWITCH

Next, we consider a more complicated case, i.e., an exclusive switch [21–23]. The exclusive switch consists of two genes, gene 1 and gene 2. The two genes have overlapping promoter sites, and the binding of one of the regulatory proteins prevents the binding of the other regulatory proteins.

Here, the interaction between the binding sites and proteins is assumed to be monomer interactions. Because the interaction is not cooperative bindings, the macroscopic rate equation gives only one solution [21–23]. Although there is only one solution, it has been shown that the exclusive switch can play as a switch. In the exclusive switch, stochastic effects make the bistability even without cooperativity between the regulatory proteins.

In order to study the exclusive switch analytically, master equations for a joint probability  $P(n_1, n_2, s_1, s_2)$  should be constructed;  $n_i \in \mathbf{N}$  is the number of proteins

for gene  $i$ , and  $s_i \in \{\text{ON}, \text{OFF}\}$  indicates the gene state. In general, the master equations for multiple gene cases are very complicated, and it could be difficult to obtain numerical solutions if the number of genes is large.

However, in our approximation scheme, the effective interaction is used, and it enables us to reduce the equations to be solved. Because we consider only the effective interaction, the probability  $P(n_1, n_2, s_1, s_2)$  can be expressed as  $P(n_1, s_1) \times P(n_2, s_2)$ . For example, the master equation for gene 1 is written as

$$\frac{d\alpha_n^{(1)}}{dt} = g^{(1)}[\alpha_{n-1}^{(1)} - \alpha_n^{(1)}] + k^{(1)}[(n+1)\alpha_{n+1}^{(1)} - n\alpha_n^{(1)}] - \tilde{h}^{(1)}\alpha_n^{(1)} + f^{(1)}\beta_n^{(1)}, \quad (29)$$

$$\frac{d\beta_n^{(1)}}{dt} = +k^{(1)}[(n+1)\beta_{n+1}^{(1)} - n\beta_n^{(1)}] + \tilde{h}^{(1)}\alpha_n^{(1)} - f^{(1)}\beta_n^{(1)}, \quad (30)$$

where the super-script ‘(1)’ indicates gene 1. Master equations for gene 2 can be obtained in the similar way. Using the same discussion in the self-regulating systems in Sec. III, the generating functions for genes  $i \in \{1, 2\}$ ,  $\alpha^{(i)}(z)$  and  $\beta^{(i)}(z)$ , are derived.

The effective interaction  $\tilde{h}^{(1)}$  should be chosen as follows. The transition of gene 1 from ON state to OFF state can occur only when the gene 2 is in ON state, and the effective interaction  $\tilde{h}^{(1)}$  includes only a contribution from the free proteins 2 in ON state. Note that gene 1 does not know whether gene 2 is in ON state or OFF state, different from the self-regulating system discussed in Sec. III; in the self-regulating system, the gene knows the own state. Hence, the evaluation of the effective interaction is slightly different from the self-regulating systems. The conditional average of the number of free proteins 2 is given by  $\langle n^{(2)} \rangle_{\alpha'} = (\partial \alpha^{(2)}(z) / \partial z|_{z=1}) / \alpha^{(2)}(1)$ , as discussed in Sec. III C. The probability  $P(2_{\text{ON}})$ , with which gene 2 is in ON state, is calculated as  $P(2_{\text{ON}}) = \alpha^{(2)}(1)$ . Defining  $\langle n^{(2)} \rangle'_{\alpha'} \equiv \langle n^{(2)} \rangle_{\alpha'} P(2_{\text{ON}})$ , the effective interaction should be written as

$$\tilde{h}^{(1)} = h^{(1)} \langle n^{(2)} \rangle'_{\alpha'}. \quad (31)$$

According to the above discussions, we finally obtain the following self-consistent equations;

$$h^{(2)} \langle n^{(1)} \rangle'_{\alpha'} = \frac{h^{(2)} g^{(1)} f^{(1)} (k^{(1)} + f^{(1)})}{k^{(1)} (f^{(1)} + h^{(1)} \langle n^{(2)} \rangle'_{\alpha'}) (k^{(1)} + f^{(1)} + h^{(1)} \langle n^{(2)} \rangle'_{\alpha'})}, \quad (32)$$

$$h^{(1)} \langle n^{(2)} \rangle'_{\alpha'} = \frac{h^{(1)} g^{(2)} f^{(2)} (k^{(2)} + f^{(2)})}{k^{(2)} (f^{(2)} + h^{(2)} \langle n^{(1)} \rangle'_{\alpha'}) (k^{(2)} + f^{(2)} + h^{(2)} \langle n^{(1)} \rangle'_{\alpha'})}. \quad (33)$$

By solving the above self-consistent equations, we obtain  $\tilde{h}^{(1)}$  and  $\tilde{h}^{(2)}$ . We here solve them numerically.

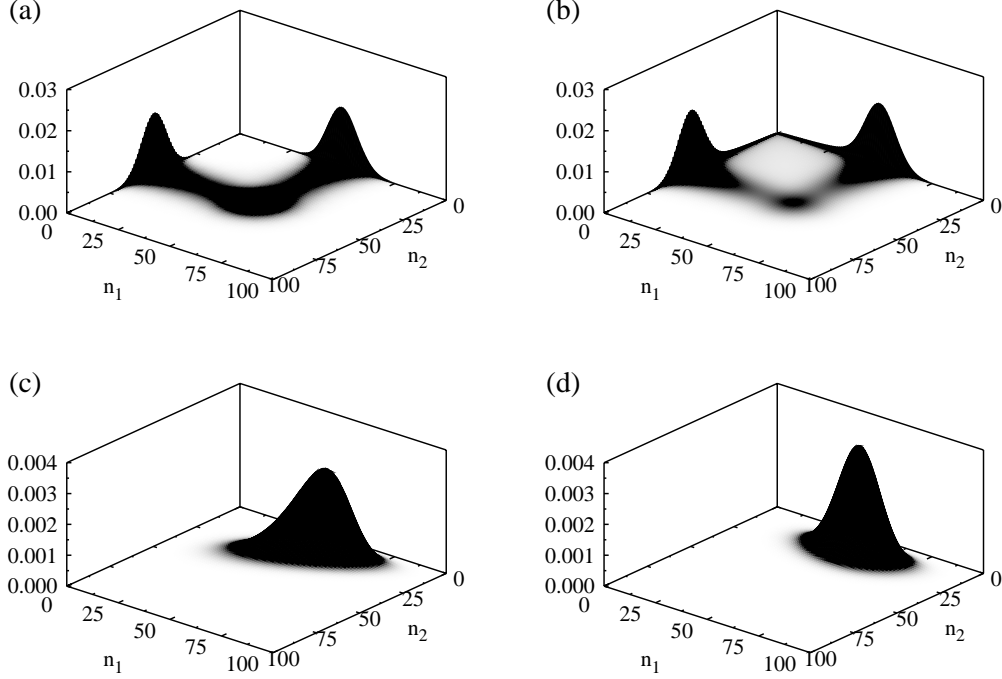


FIG. 7: Joint probability distributions for the exclusive switch. (a) Monte Carlo results for  $X_{\text{eq}}^{(1)} = X_{\text{eq}}^{(2)} = 25.0$ ,  $X_{\text{ad}}^{(1)} = X_{\text{ad}}^{(2)} = 25.0$ , and  $\omega^{(1)} = \omega^{(2)} = 0.1$ . (b) Approximate results. All parameters are the same as (a). (c) Monte Carlo results for  $X_{\text{eq}}^{(1)} = 40.0$ ,  $X_{\text{eq}}^{(2)} = 30.0$ ,  $X_{\text{ad}}^{(1)} = 30.0$ ,  $X_{\text{ad}}^{(2)} = 20.0$ ,  $\omega^{(1)} = 10.0$  and  $\omega^{(2)} = 20.0$ . (d) Approximate results. All parameters are the same as (c).

We can immediately obtain the probability distributions for each gene using the approximation scheme. In order to reconstruct the joint probability distribution for genes 1 and 2, we need more calculations as follows. Firstly, we calculate conditional probabilities for the number of free proteins for gene  $i$  ( $i \in \{1, 2\}$ ) as

$$\tilde{\alpha}'_n^{(i)} \equiv \frac{\alpha'_n^{(i)}}{\alpha'^{(i)}(1)}, \quad (34)$$

$$\tilde{\beta}_n^{(i)} \equiv \frac{\beta_n^{(i)}}{\beta^{(i)}(1)}. \quad (35)$$

Secondly, because of the approximation scheme, the joint probability distribution should be evaluated as

$$P(n_1, n_2, 1_{\text{ON}}, 2_{\text{ON}}) = P(1_{\text{ON}}, 2_{\text{ON}}) \tilde{\alpha}'_{n_1-1}^{(1)} \tilde{\alpha}'_{n_2-1}^{(2)}, \quad (36)$$

$$P(n_1, n_2, 1_{\text{ON}}, 2_{\text{OFF}}) = P(1_{\text{ON}}, 2_{\text{OFF}}) \tilde{\alpha}'_{n_1-1}^{(1)} \tilde{\beta}_{n_2}^{(2)}, \quad (37)$$

$$P(n_1, n_2, 1_{\text{OFF}}, 2_{\text{ON}}) = P(1_{\text{OFF}}, 2_{\text{ON}}) \tilde{\beta}_{n_1}^{(1)} \tilde{\alpha}'_{n_2-1}^{(2)}, \quad (38)$$

where  $P(1_{\text{ON}}, 2_{\text{ON}})$  is the probability with which gene 1 is in ON state and gene 2 is in ON state, and so on. Note that  $\alpha'$  means only the number of 'free' proteins; for monomer interaction cases, the difference between  $\alpha$  and  $\alpha'$  is only one inert protein. In addition, the probability

with which both genes 1 and 2 are in OFF state is zero;  $P(1_{\text{OFF}}, 2_{\text{OFF}}) = 0$ , because of the exclusive settings.

Taking the exclusive settings into account, the marginal probabilities are calculated as follows:

$$\begin{cases} P(1_{\text{ON}}, 2_{\text{ON}}) + P(1_{\text{OFF}}, 2_{\text{ON}}) = P(2_{\text{ON}}), \\ P(1_{\text{ON}}, 2_{\text{OFF}}) = P(2_{\text{OFF}}), \\ P(1_{\text{OFF}}, 2_{\text{ON}}) = P(1_{\text{OFF}}), \\ P(1_{\text{ON}}, 2_{\text{ON}}) + P(1_{\text{ON}}, 2_{\text{OFF}}) = P(1_{\text{ON}}), \end{cases} \quad (39)$$

and then

$$\begin{cases} P(1_{\text{ON}}, 2_{\text{ON}}) = P(1_{\text{ON}}) - P(2_{\text{OFF}}), \\ P(1_{\text{ON}}, 2_{\text{OFF}}) = P(2_{\text{OFF}}), \\ P(1_{\text{OFF}}, 2_{\text{ON}}) = P(1_{\text{OFF}}), \\ P(1_{\text{OFF}}, 2_{\text{OFF}}) = 0. \end{cases} \quad (40)$$

The marginal probabilities, such as  $P(1_{\text{ON}})$ , can be evaluated by using the generating function  $\alpha'^{(i)}(z)$  and  $\beta^{(i)}(z)$ . Finally, we can construct the joint probability distribution as

$$\begin{aligned} P(n_1, n_2) = & P(n_1, n_2, 1_{\text{ON}}, 2_{\text{ON}}) \\ & + P(n_1, n_2, 1_{\text{OFF}}, 2_{\text{ON}}) \\ & + P(n_1, n_2, 1_{\text{ON}}, 2_{\text{OFF}}). \end{aligned} \quad (41)$$

We here note that the probabilities  $P(1_{\text{ON}}, 2_{\text{ON}})$ , calculated using the above procedures, may become negative for some cases; i.e.,  $P(1_{\text{ON}}) > P(2_{\text{OFF}})$  for some choices of parameters  $g^{(i)}$ ,  $k^{(i)}$ ,  $h^{(i)}$  and  $f^{(i)}$ . In these cases, other procedures to estimate the joint probabilities are needed. In the following numerical experiments, only the former cases ( $P(1_{\text{ON}}) < P(2_{\text{OFF}})$ ) are treated.

Figure 7 shows the joint probability distributions. Figures 7(a) and (c) are Monte Carlo results, and Figs. 7(b) and (d) are results of the approximation scheme. As in Sec. IIID, we used the rescaled parameters, and set  $k^{(1)} = k^{(2)} = 1$ . In Figs. 7(a) and (b), we used the parameters  $X_{\text{eq}}^{(1)} = X_{\text{eq}}^{(2)} = 25.0$ ,  $X_{\text{ad}}^{(1)} = X_{\text{ad}}^{(2)} = 25.0$ , and  $\omega^{(1)} = \omega^{(2)} = 0.1$ ; for (c) and (d),  $X_{\text{eq}}^{(1)} = 40.0$ ,  $X_{\text{eq}}^{(2)} = 30.0$ ,  $X_{\text{ad}}^{(1)} = 30.0$ ,  $X_{\text{ad}}^{(2)} = 20.0$ ,  $\omega^{(1)} = 10.0$  and  $\omega^{(2)} = 20.0$ . The numbers of the Monte Carlo steps are over  $10^8$  for Fig. 7(a), and over  $10^9$  for Fig. 7(c). Although Fig. 7(d) does not show the correlated behavior seen in Fig. 7(c) because correlations between gene 1 and gene 2 are largely neglected in the approximation scheme, one could say that the approximation scheme gives qualitatively good results; the characteristics of the peak structure can be recovered adequately despite the rough approximation. In Fig. 7(b), the bistability due to the exclusive settings is recovered well.

## V. CONCLUDING REMARKS

In the present paper, we developed the approximation scheme for gene regulatory systems. We firstly applied it to self-regulating systems. The approximation scheme gives qualitatively good results; the characteristics of peak structures can be recovered well. In addition,

due to the extension of the basic idea of the effective interactions, we can naturally apply the approximation scheme even to the exclusive switch, and the bistability of the exclusive switch without cooperative interactions is successfully recovered.

In contrast to the Fokker-Planck or Langevin approach, the approximation scheme proposed in the present paper does not neglect discrete properties of systems. In addition, because we can rewritten the joint probability for all genes as a product of probability distribution for each gene, the dimensions of the problems are reduced largely.

The approximation scheme developed in the present paper would be a very crude one; it cannot treat correlated characteristics between genes. However, approximate analytic expressions are immediately obtained, and qualitatively good results are given despite the crude approximation. Since Monte Carlo simulations are sometimes time-consuming and need high computational costs, it would be beneficial to study such approximation scheme in order to obtain qualitative pictures for the probability distributions. In addition, developments of analytical treatments would be helpful to gain insights and understandings for the regulatory systems

## ACKNOWLEDGMENTS

The author thank Masaki Sasai for helpful comments for this manuscript. This work was supported in part by grant-in-aid for scientific research (Nos. 20115009 and 21740283) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

- 
- [1] C. V. Rao, D. M. Wolf, and A. P. Arkin, *Nature* **420**, 231 (2002).
  - [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
  - [3] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, *Nature Rev. Genetics* **6**, 451 (2005).
  - [4] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature* **403**, 339 (2000).
  - [5] H. Okano, T. J. Kobayashi, H. Tozaki, and H. Kimura, *Biophys. J.* **95**, 1063 (2008).
  - [6] D. Schultz, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **126**, 245102 (2007).
  - [7] M. Yoda, T. Ushikubo, W. Inoue, and M. Sasai, *J. Chem. Phys.* **126**, 115101 (2007).
  - [8] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, *Proc. Natl. Acad. Sci U.S.A* **97**, 2075 (2000).
  - [9] T. B. Kepler and T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
  - [10] W. Bialek, *Adv. Neural Inf. Proc.* **13**, 103 (2001).
  - [11] M. Sasai and P. G. Wolynes, *Proc. Natl. Acad. Sci U.S.A* **100**, 2374 (2003).
  - [12] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, *Phys. Rev. E* **72**, 051907 (2005).
  - [13] A. M. Walczak, M. Sasai, and P. G. Wolynes, *Biophys. J.* **88**, 828 (2005).
  - [14] K.-Y. Kim and J. Wang, *PLoS Comp. Biol.* **3**, e60 (2007).
  - [15] J. Ohkubo, *J. Stat. Mech.* P09017 (2007).
  - [16] V. Shahrezaei and P. S. Swain, *Proc. Natl. Acad. Sci U.S.A* **105**, 17256 (2008).
  - [17] P. Visco, R. J. Allen, and M. R. Evans, *Phys. Rev. E* **79**, 031923 (2009).
  - [18] A. M. Walczak and P. G. Wolynes, *Biophys. J.* **96**, 4525 (2009).
  - [19] J. L. Cherry and F. R. Adler, *J. Theor. Biol.* **203**, 117 (2000).
  - [20] P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.* **92**, 128101 (2004).
  - [21] A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, *Phys. Rev. Lett.* **96**, 188101 (2006).
  - [22] A. Loinger, A. Lipshtat, N. Q. Balaban, and O. Biham, *Phys. Rev. E* **75**, 021904 (2007).



- [23] D. Schultz, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci U.S.A **105**, 19165 (2008).
- [24] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).